DYNAMICAL TOMOGRAPHY OF GRAVITATIONALLY BOUND SYSTEMS

Mikko Kaasalainen

Department of Mathematics and Statistics
P.O. Box 68, FI-00014 University of Helsinki
Finland

(Communicated by Jari Kaipio)

Abstract. We study the inverse problem of deducing the dynamical characteristics (such as the potential field) of large systems from kinematic observations. We show that, for a class of steady-state systems, the solution is unique even with fragmentary data, dark matter, or selection (bias) functions. Using spherically symmetric models for simulations, we investigate solution convergence and the roles of data noise and regularization in the inverse problem. We also present a method, analogous to tomography, for comparing the observed data with a model probability distribution such that the latter can be determined.

1. Introduction

Matter in the universe is usually contained in systems bound together by gravitation: planets and their moons, planetary systems, star clusters, galaxies, and groups of galaxies. Indeed, the modern concepts of gravitation and gravitational potential were brought about by the realization that a mathematically well definable universal force field must keep celestial bodies in their observed orbits. Newton's solution to the inverse problem of "What kind of a force keeps two pointlike bodies in an elliptic orbit around each other?" was the inverse-square law of attraction[1]. In such a force field, the equations of motion for a body can be written as

$$\frac{d^2x}{dt^2} = -\nabla\Phi(x),$$

where $t \in \mathbb{R}$ is the time, $x \in \mathbb{R}^3$ describes the position of the body, and $\Phi : \mathbb{R}^3 \to \mathbb{R}$ is the gravitational potential; here $\Phi(x) \sim \|x\|^{-1}$.

Outside isolated two-body configurations, the two-body solution only serves as a useful first approximation for systems dominated by one massive body; for such a class of systems, the perturbations from this solution can be examined analytically

[1]Newton actually solved the direct problem of "What kind of an orbit is produced by the inverse-square attraction?" (this is what Halley asked him). The solution of the inverse problem is an instant corollary; it is also unique for motion around the focal point of an ellipse – the harmonic oscillator creates an elliptic orbit around the centre.

to a relatively high precision. The behaviour of a system consisting of a moderate number of bodies of more or less equal weight must be numerically integrated assuming the Newtonian potential for each pointlike mass. However, if the system consists of many ($> 10^6$ or so) gravitationally interacting bodies collectively forming the potential, we can assume it to obey some principles of statistical mechanics. Accordingly, we no longer analyze the motion of separate particles, but want to define the dynamical characteristics of the system as a whole, such as its global-scale potential field $\Phi(x)$, matter density $\rho(x)$, etc. We define *dynamical tomography* as the inverse problem of deducing these smoothly distributed characteristics from kinematic observations. A more detailed definition of this large-scale generalization of Newton's dynamical inverse problem is given in Section 2.1. The term tomography can be used here in both intuitive and technical senses: we determine density/distribution functions inside a domain, and this can be done by employing the principles of tomographic analysis.

Large gravitationally bound systems have been studied for more than a century, and the corresponding theoretical framework is described in several textbooks, of which [2] is the most up-to-date one. However, the studies have mostly concentrated on the direct problem of creating numerically or analytically well tractable dynamical models, or of designing approximate system models to explain various observed phenomena, while the general inverse problem (as posed in sections 2 and 3 here) has received less attention. This is mostly due to the lack of data sufficient for a comprehensive analysis, and the scarcity of efficient methods of dealing with such data. New sky surveys such as LSST (Large Synoptic Survey Telescope, USA) and Gaia (EU), both starting at the beginning of the next decade, will provide a vast amount of kinematic data of the stars in our galaxy, making the construction of a robust machinery for inverse problem analysis essential. We describe here some theoretical and practical aspects of solving problems in dynamical tomography.

The structure of the paper is as follows: in Section 2, we review as well as define some basic concepts of large gravitational systems, and pose the problem in corresponding terminology. In Section 3, we examine the fundamental aspects and uniqueness properties of the inverse problem, and in Section 4 present examples in the simplified case of spherically symmetric systems. In Section 5, we redefine the problem in the sense of probability distributions, which is usually necessary in practice. Finally, in Section 6, we present conclusions and discuss future work and applications.

2. Defining large systems

2.1. **Distribution functions.** In this paper, we assume the system to be bound, stable, collisionless ($d \gg D$, where d is the shortest distance between two bodies of size D) and in a steady-state equilibrium (for a discussion on these standard assumptions and their applicability regimes, see [2]). In practice, this is a good approximation for large stellar systems, much in the same way as the two-body solution is a good starting point for analyzing the solar system. The dynamics of such a system is completely described by the smooth phase-space distribution function $f(x, v, t)$ of its constituent particles (stars), $f : \mathbb{R}^3_x \times \mathbb{R}^3_v \times \mathbb{R}_t \to \mathbb{R}$, where $v \in \mathbb{R}^3$ is the velocity. The function f pertains to the limits $d \to 0$ (infinitely many particles), $D \to 0$, and $D/d \to 0$. In practice, f is interpreted as the quantity giving the expected number of particles in a phase-space volume via $f \, d^3x \, d^3v$, or as the probability density of finding a particle at (x, v).

The distribution function f essentially corresponds to fluid in six dimensions; the equation of motion for f is the collisionless Boltzmann (or Vlasov) equation

$$(1) \qquad \frac{\partial f}{\partial t} + \langle v, \nabla_x f \rangle - \langle \nabla \Phi, \nabla_v f \rangle = 0$$

(a special case of Liouville's theorem; see [2]). In terms of Poisson brackets, this reads $\partial f / \partial t = [H, f]$, where H is the Hamiltonian. This implies that the flow of the fluid (the smooth motion of particle phase points through phase space) is incompressible. In this paper, we study the steady-state case where matter units move in their orbits in the potential field Φ while leaving the collective system configuration completely unchanged: $\partial f / \partial t = 0 = \partial \Phi / \partial t$, i.e., $f = f(x, v)$ and $\Phi = \Phi(x)$. Moreover, the observational data, measurements of (x, v) for stars or equivalent units of matter, are effectively for one epoch: though they are measured over several years, the orbital periods are millions of years, so the data do not provide orbital information outside the linearized regime $v = dx/dt$.

We also define a set Q including all the other characteristics than x, v that can be assigned to a star from observations, such as luminosity, temperature, chemical composition, etc. These may be direct observables or implicit quantities such as mass and age that can be expressed in terms of direct ones. They can be continuous or even discrete (taxonomy), in which case an integral over such a variable is understood in a suitable corresponding sense. Thus we have a full distribution function $F(x, v; Q)$ of $N+6$ dimensions, where N is the number of the quantities in the set Q. At first sight, one might imagine that only mass could be a dynamically interesting quantity in Q, but, e.g., luminosity directly affects the observability and bias factors. Also, as will be shown below, the possibility to partition the observed particles into different populations with the aid of Q adds to the information content.

We can distribute $f(x, v)$ further among the members of Q at each (x, v) with a function $\varphi(x, v; Q)$:

Definition 2.1. The full distribution function of number density in $\mathbb{R}_x^3 \times \mathbb{R}_v^3$ and in the set Q of N quantities q (generic notation for a member of Q) is

$$(2) \qquad F(x, v; Q) = f(x, v)\varphi(x, v; Q),$$

normalized such that $\int_Q \varphi(x, v; Q) \, d^N q = 1$. The marginal distribution of some $q \in Q$, $q_0 \le q \le q_1$ is

$$(3) \qquad F(x, v; q) = f(x, v)\varphi(x, v; q) = f(x, v) \int_{Q \setminus q} \varphi(x, v; Q) \, d^{N-1} q'$$

and $\int_{q_0}^{q_1} \varphi(x, v; q) \, dq = 1$.

Definition 2.2. The distribution function $f_i(x, v)$ of an object population P_i is given by

$$(4) \qquad f_i(x, v) = f(x, v)\mathcal{P}_i(x, v),$$

where the population filter $\mathcal{P}_i(x, v)$ is

$$(5) \qquad \mathcal{P}_i(x, v) = \int_{\Lambda_i} \varphi(x, v; Q) \, d^N q,$$

where Λ_i sets the (integration) limits for each observable of Q according to the given characteristics of P_i.

An important distribution function is that of mass per $d^3x\, d^3v$, denoted by $\tilde{f}(x,v)$:

Definition 2.3. The mass distribution function is given by

$$\tilde{f}(x,v) = f(x,v)\mathcal{M}(x,v), \tag{6}$$

where the mass factor $\mathcal{M}(x,v)$ is

$$\mathcal{M}(x,v) = \int_0^\infty M\varphi(x,v;M)\, dM. \tag{7}$$

For example, the case of all particles having the same mass M_0 is given by

$$\varphi(x,v;M) = \delta(M - M_0),$$

where δ is the Dirac delta distribution. Thus $\mathcal{M} = M_0$.

The amount of matter in an infinitesimal volume of phase space $\mathbb{R}_x^3 \times \mathbb{R}_v^3$ is $\tilde{f}\, d^3x\, d^3v$, and the matter density $\rho(x)$ in $\mathbb{R}_x^3$ is

$$\rho(x) = \int_{V_+} \tilde{f}(x,v)\, d^3v, \tag{8}$$

where the velocity domain V_+ includes all velocities v that can exist at x in a bound system; alternatively, for $v \notin V_+$ we define $f(x,v) = 0$ and can integrate over all v. The matter density and the potential are related through Poisson's equation:

$$\rho(x) = \frac{1}{4\pi G}\nabla^2 \Phi(x), \tag{9}$$

where G is the universal gravitation constant. Note that if $\tilde{f}$ is the distribution of all matter, equations (1), (8) and (9) must be simultaneously fulfilled. The search for such self-consistent solution pairs of $\tilde{f}, \Phi$ is the fundamental problem of stellar dynamics. However, as we will show in Section 3, f (or $\tilde{f}$) does not need to cover all matter in our problem; it can describe a selected population, or there may be unobservable matter such that the full f cannot be determined while Φ can.

The distribution function f is a statistical tool, meant to describe the average distribution of matter in large enough bins. As mentioned above, another useful practical interpretation of f is the probability density of observing a star (or, for $F(x,v;Q)$, a star with properties Q) at (x,v); this circumvents the local-scale distribution of matter in space, and the size of our averaging bins – the collisionless approximation means (and observations show) that most of space is void of luminous matter. In section 5, we will use this interpretation to redefine the observables of our inverse problem.

We can now state our problem of dynamical tomography as follows:

Problem. *Given a large number of observed (x,v) (for any motion markers such as stars or other matter and possibly in different populations P_i) in a domain $\Omega \subset \mathbb{R}^3 \times \mathbb{R}^3$ in a gravitationally bound steady-state system, deduce the potential $\Phi(x)$ of the system, and the distribution function(s) $f(x,v)$ of the observed matter in Ω.*

Obviously, if the number of observations is extremely large and all matter in the system is observed, the problem is trivial. The number of observations in arbitrarily small phase-space volumes is now nonzero, so an accurate estimate of the distribution $f(x,v)$ is obtained by simply counting the objects, assuming that their masses are known. From this, $\Phi(x)$ follows; one could, as a shortcut to $\Phi(x)$, directly count objects in $\mathbb{R}_x^3$ only. What makes the problem an inverse one is that

neither of the requisites is fulfilled in reality: the number of observations, though high, is not sufficient for direct counting in small enough phase-space (or even $\mathbb{R}^3_x$) bins, and, above all, not all of the system matter is observed because of bias factors or non-luminous matter. This is why we need to consider $\mathbb{R}^3_x \times \mathbb{R}^3_v$ instead of just $\mathbb{R}^3_x$ even if we only want to obtain $\Phi(x)$: the fraction of unobserved matter can only be inferred by using a priori information furnished by the expected dynamical properties of the system. Our goal is now to find a parametrization in which the dynamics and a priori constraints of the system can be expressed in a practical and preferably analytically tractable way. Here we neglect effects such as spatial correlation between stars, changing mass (shedding or accretion) due to stellar evolution, etc. [2].

Some basic properties of the system can already be deduced from a very small number of observations. For example, if we observe matter at distance r from the centre of the system, moving at speed v_r relative to us, and assume it to be bound by the system, we can roughly estimate the mass M inside r, e.g., by assuming a spherical mass distribution and a circular orbit around the centre: $M = v_r^2 r / G$. In this and similar ways it has been deduced from several observations of galaxies and their clusters that most of the mass in the universe is contained in dark matter: non-luminous and extremely hard if not impossible to observe directly.

2.2. JEANS THEOREMS AND INTEGRABLE SYSTEMS. As is well known, a natural and compact way of parametrizing the desired dynamical properties of the system can be achieved via its dynamical invariants. A function $I(x,v)$, evaluated at $[x(t), v(t)]$ of any orbit in the potential Φ of the system, is an *integral of motion* if

$$\frac{d}{dt} I[x(t), v(t)] = 0$$

for all orbits at all t. Evaluating the time derivative by the chain rule and substituting the equation of motion (1), we obtain an equation exactly of the form (1) (I corresponding to f) with the steady-state condition $\partial f / \partial t = 0$. Thus we arrive at the Jeans theorem (see, e.g., [2]) that states that any steady-state solution f of (1) is of the form $f[I_i(x,v)]$, where I_i are integrals of motion. For example, the energy $E(x,v)$:

$$(10) \qquad\qquad E(x,v) = \frac{1}{2} \langle v, v \rangle + \Phi(x)$$

is always an integral in gravitationally interacting systems.

To have a complete set of integrals which we can, in principle, reconstruct and thus properly use in our formulation of the problem, we further restrict ourselves to regular orbits and integrable systems (defined below), a subset of the steady-state solutions of (1). This class of solutions, though restricted, has two great advantages:

  i) it facilitates an analytical or semianalytical investigation of the problem as well as the derivation of uniqueness results; and
 ii) it can be readily viewed as a further modifiable approximate representation of the full class of solutions.

An orbit is regular (quasiperiodic) if it is confined to a 3-torus $S^1 \times S^1 \times S^1$ in phase space [1]. Then it has three integrals (also called isolating integrals) $I_i, i = 1, 2, 3$ that can be used to define the torus: $I_i(x,v) = C_i$, where C_i are constants. The isolating integrals can be given in an arbitrary basis as functions of I_i are isolating integrals as well. A proper basis set of isolating integrals maps each torus

$\mathcal{T}$ as $\mathcal{T} \leftrightarrow I_i, i = 1, 2, 3$, i.e., any such bases can be mapped one-to-one into each other: $\{I_i'\} \leftrightarrow \{I_i\}$.

If a potential $\Phi(x), x \in \mathbb{R}^3$ creates an integrable system, all orbits are confined to 3-tori in $\mathbb{R}_x^3 \times \mathbb{R}_v^3$ that are each defined by three action (Poincaré) integrals $J_i, i = 1, 2, 3$, a class of isolating integrals I:

$$(11) \qquad\qquad J_i = \frac{1}{2\pi} \oint_{\mathfrak{P}_i} \langle p, dq \rangle,$$

where $p \in \mathbb{R}^3$ and $q \in \mathbb{R}^3$ are any canonically conjugate momenta and coordinates, and $\mathfrak{P}_i$ is a path that cannot be continuously deformed into a point. For other paths, the integral vanishes. There are three such sets of possible paths each producing one J_i; see, e.g., [1, 2]. $J \in \mathbb{R}_J^3$ plays the role of canonical momentum in Hamilton's equations, and its canonically conjugate coordinate pair is the angle variable $\theta \in \mathbb{R}^3$. (J, θ) are related to phase-space coordinates (x, v) by a canonical transformation, and energy (Hamiltonian) depends on J only: $E = E(J)$.

Action-angle formalism yields the time averages theorem [2] that states that the density of the orbit on the torus, i.e., the average time it spends on different parts of the torus, is evenly distributed in θ on its surface. This holds for non-resonant, i.e., not closed orbits that make up almost all of phase space. This, in turn, leads to the strong Jeans theorem that essentially states that, in an integrable system, a steady-state solution of (1) must be of the form

$$(12) \qquad\qquad f(x, v) = f[I_1(x, v), \dots, I_3(x, v)] := f[I(x, v)],$$

where I_i are isolating integrals.

If we want a steady-state solution for the luminous matter in the system, the distribution function of our choice is the mass distribution $\tilde{f}(x, v)$, i.e., we now assume that $\tilde{f} = \tilde{f}[I(x, v)]$. Thus, if we want f and $\tilde{f}$ to be interchangeable in our steady-state equations and theorems to be derived in section 3, we must assume the mass factor $\mathcal{M}$ to be of the form $\mathcal{M} = \mathcal{M}[I(x, v)]$ as well, i.e., $\varphi(x, v; M) = \varphi[I(x, v); M]$. Leaving the masses unknown is the advantage in using f, but in practice one should solve the inverse problem with $\tilde{f}$ as well by weighting the count of each observed point (x_i, v_i) by the corresponding mass M_i, and then compare the results (e.g., the tori obtained).

The fact that the system as a whole is a steady-state and integrable one does not generally imply that a population filter $\mathcal{P}_i(x, v)$ is of the form $\mathcal{P}_i[I(x, v)]$, or that, for any other q than the mass M, $\varphi(x, v; q) = \varphi[I(x, v); q]$. But if such filters (i.e., steady-state populations) can be assumed to be identifiable at least in some domain of Q and $\mathbb{R}_x^3 \times \mathbb{R}_v^3$, they are very useful in the inverse problem, as will be shown in section 3.

3. Inverse problem

3.1. Uniqueness. By virtue of the strong Jeans theorem, we now seek integrable systems that reproduce the observations. Since the orbits of all particles, observed or not, are on tori, and $f(x, v)$ is of the form (12), we know that each isosurface $f(x, v) = const$ in phase space $\mathbb{R}_x^3 \times \mathbb{R}_v^3$ of any steady-state $f(x, v)$, regardless of the group of objects it represents, entirely consists of 3-tori on which $I(x, v) = const$. To show that these isosurfaces determine the potential $\Phi(x)$ of the system, let us first present the following lemma:

Lemma 3.1. *Let the potential $\Phi(x)$ generate an integrable system, and let $\mathcal{T}$ denote the corresponding set of 3-tori in $\mathbb{R}^3 \times \mathbb{R}^3$. Then $\Phi(x)$ is the only integrable potential (up to an additive constant) that creates any chosen subset $\widehat{\mathcal{T}}$ of arbitrarily small patches Γ on any tori of $\mathcal{T}$ such that $\widehat{\mathcal{T}}$ covers all of $\mathbb{R}^3_x$ (accessible to the system) in a connected manner.*

Proof. The value E of the energy (10) is constant on each torus. Let us choose a patch Γ_0 on which $\Phi(x) = E_0 - \frac{1}{2}\langle v, v \rangle$, where E_0 is the energy on the corresponding torus. Thus $\Phi(x)$ is defined up to an arbitrary constant for all x of the torus. If we have another patch Γ_1 with a point $x = x_0$ common with Γ_0, and the corresponding phase-space points on the two patches are, respectively, (x_0, v_0) and (x_0, v'), the value E_1 of the energy on Γ_1 must be

$$E_1 = \Phi(x_0) + \frac{1}{2}\langle v', v' \rangle = E_0 + \frac{1}{2}(\langle v', v' \rangle - \langle v_0, v_0 \rangle).$$

This defines $\Phi(x)$ on Γ_1: $\Phi(x) = E_1 - \frac{1}{2}\langle v, v \rangle$. This chain of patches can be continued to cover $\Phi(x)$ everywhere (up to the arbitrary constant chosen for E_0 at the beginning). $\square$

Remark 1. The connected manner thus means that Γ on different tori can be arranged in a sequence by common values of x as above. Actually, any patch on a torus can even be replaced by a set of two points that do not have to be close to each other: to establish the connected chain for $\Phi(x)$, we only require two points with different x on one torus, and one of the x shared with a point on another torus so that all x are covered.

Remark 2. The lemma can also be expanded to concern all potentials (not just integrable systems) by defining Γ to be sections of orbits having common points x. In fact, just one chaotic orbit is sufficient as it eventually defines $\Phi(x)$ at all $x \in \mathbb{R}^3$. More generally, Γ can denote parts of any structures on which E is constant, or parts of isosurfaces of any functions of the form $f(E)$. Note that an isosurface need not be connected, i.e., $f(E)$ need not be monotonous. One value of f can correspond to more than one value of E as long as the branches of different E can be identified: $\Gamma \leftrightarrow E$.

The lemma states that even highly fragmentary information on the shape of the tori in phase space is well sufficient to determine the integrable potential $\Phi(x)$ uniquely. Now, let independent distribution functions $f_i(x, v)$, $i = 1, 2, 3$, in an integrable potential $\Phi(x)$ be defined everywhere in $\mathbb{R}^3 \times \mathbb{R}^3$. By the strong Jeans theorem, the isosurfaces $f_i(x, v) = const$ are of the form $f[I(x, v)] = const$, where $I(x, v) = const$ define the tori created by $\Phi(x)$. The 3-surfaces formed by the intersection of three 5-surfaces $f_i(I)$ are 3-tori (defined by I as well). Then, by the lemma, any collection of parts of surfaces $f_i = const$ sufficient to determine a connected chain of torus patches uniquely determines $\Phi(x)$. We can now state the following uniqueness theorem:

Theorem 3.2. *Let three independent steady-state distribution functions $f_i(x, v)$, $i = 1, 2, 3$ ($f : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$) of matter in an integrable system be defined in some common domains of $\mathbb{R}^3 \times \mathbb{R}^3$ such that a set of parts of the surfaces $f_i(x, v) = const$ forms a succession of torus patches connected in x. Then the $f_i(x, v)$ uniquely determine the potential $\Phi(x)$.*

Remark 3. We can pose this in the integral space $\mathbb{R}_I^3$ as well: we assume that the $f_i(I)$, $I \in \mathbb{R}^3$, are such that the set of equations $\{f_i(I) = C_i\}$ has a nonzero and finite number of solutions I for sufficiently many $\{C_i\}$ such that a connected sequence of patches can be constructed (both I and $f_i(I)$ are, of course, unknown prior to the determination of $\Phi(x)$). Note that the number can be larger than one, i.e., $f_i(x,v)$ need not form a proper basis of isolating integrals: if a set $\{f_i = C_i\}$ corresponds to more than one torus (more than one set of $\{I_i = C_i'\}$), it is sufficient to be able to distinguish the different tori in the sense of the lemma (cf. remark 2 of the lemma).

Remark 4. The theorem emphasizes the role of different distribution functions $f_i(x,v)$ in providing information about the system. If different f_i based on the observations of various steady-state object populations are available, we gain more by studying f_i rather than the total distribution function $f_{\text{tot}} = \sum_i f_i$ of all luminous mass.

Remark 5. Most of all the feasible ways of filling $\mathbb{R}^3 \times \mathbb{R}^3$ with invariant tori, each with a constant J of (11), are not derivable from any potential (cf. [7]). Thus there usually is no integrable potential ϕ other than Φ that, with its tori I_ϕ and some distribution function f', would yield $f'[I_\phi(x,v)] = f[I_\Phi(x,v)]$ everywhere in $\mathbb{R}^3 \times \mathbb{R}^3$, so in most cases one distribution function $f(x,v)$ in $\mathbb{R}^3 \times \mathbb{R}^3$ determines $\Phi(x)$.

The theorem is constructive for three f's from which we directly get $\Phi(x)$ with the procedure of the lemma, while the case of one f (remark 5) is non-constructive. Finding the integrable $\Phi(x)$ corresponding to one f is not obvious since there are no general procedures for finding and exploring integrable potentials [10]. The set of potentials giving rise to integrable systems is known to be a vanishingly small subset of all potentials [10], and there is no such $\epsilon \neq 0$ that, for an integrable $\phi(x)$, $\phi(x) + \epsilon\varphi(x)$ is still integrable for an arbitrary $\varphi(x)$. In practice, we can circumvent this difficulty by allowing the use of non-integrable potentials and approximate tori, i.e., we construct a set of approximate integrals $I(x,v)$ for a potential $\Phi(x)$ [6]; this approach will be discussed in section 6.

The importance of isosurfaces in an integrable system is that they relieve us from having to record the time evolution of orbits. If the orbital tori are definable, we only need to probe phase space at one moment. It is important to note that $f(x,v)$ does not need to cover all matter in the system (then the single-f uniqueness would be trivial for all systems): it can represent any fraction of the matter, even virtually massless test particles, as long as it is defined on all tori and it has settled to a steady-state equilibrium. This underlines the fact that we are interested more in the geometric structures in $\mathbb{R}^3 \times \mathbb{R}^3$ (in practice, in a sufficiently large domain Ω of observations) rather than the actual distribution function f itself.

The fact that we do not need to observe all of the matter in the system to determine $\Phi(x)$ is of crucial importance. Because of the abundance of dark matter, a large part of all matter will not be seen even in ideal conditions. If, for example, f represents all luminous matter, we immediately obtain the mass density $\rho_D(x)$ of dark matter by subtracting (8) from (9) :

$$(13) \qquad \rho_D(x) = \frac{1}{4\pi G}\nabla^2\Phi(x) - \int_{V_+} \tilde{f}(x,v)\, d^3v.$$

Another cause of unobservability is the inevitable fact that our instruments are not sensitive enough, or that, e.g., interstellar dust prevents us from seeing some regions of space properly. Such effects can be described by the bias or selection function $\gamma(x, v)$, $0 < \gamma \leq 1$, that represents the fraction of observable matter of $f(x, v)$ in $d^3x\, d^3v$. The potential $\Phi(x)$ can be determined uniquely even in the presence of $\gamma(x, v)$ as we immediately can see if we suppose that we have independent $f_i(x, v) > 0$ available that share the same $\gamma(x, v)$, since now the ratios of f_i can be viewed as a new basis of isolating integrals:

Theorem 3.3. *Let the products $g_i(x, v) = \gamma(x, v)\, f_i(x, v) > 0$, $i = 1, \ldots, 4$, of a bias function $0 < \gamma \leq 1$ and four independent steady-state distribution functions f_i be defined in $\mathbb{R}^3 \times \mathbb{R}^3$. Then the three ratios*

$$\frac{g_i(x, v)}{g_1(x, v)} = \frac{f_i(x, v)}{f_1(x, v)}, \quad i = 2, 3, 4,$$

uniquely determine the potential $\Phi(x)$ as in theorem 3.2.

Remark 6. As above, this emphasizes the combined information content of different distribution functions. Analogously with remark 5, we can stipulate for two f_i that in most cases two products $g_i(x, v) = \gamma(x, v)\, f_i(x, v) > 0$, $i = 1, 2$, defined everywhere in $\mathbb{R}^3 \times \mathbb{R}^3$, determine $\Phi(x)$ via their ratio $g_2(x, v)/g_1(x, v) = f_2(x, v)/f_1(x, v)$.

Remark 7. The bias function $\gamma(x)$ may depend on x only; $\gamma : \mathbb{R}^3_x \to \mathbb{R}$. In such cases the additional factor $\gamma(x)$ does not really increase the possibility of the existence of another integrable potential $\phi(x)$ and bias function $\xi(x)$, $0 < \xi \leq 1$, such that $\xi f' = \gamma f$ everywhere. This would require there to be an isolating integral of $\phi(x)$ that everywhere matches an isolating integral of another potential ($\Phi(x)$) when multiplied with a function of x only ($\xi(x)/\gamma(x)$), while general isolating integrals are mixed functions of x, v, and individual potentials. No examples of such transformations are known. Thus, usually the single product $\gamma(x)f(x, v)$ determines the potential $\Phi(x)$ of the system.

The above results mean that dark matter, selection bias, or fragmentary data are no fundamental obstacles to dynamical tomography. It should be noted that, for near-integrable systems in the sense of Kolmogorov-Arnold-Moser (KAM) theorem [1], most of our results for integrable potentials still hold to a high precision. This is because they are based on the toroidal topology of motion in phase space, and almost all motion in near-integrable systems still occurs on tori that can be viewed as perturbed versions of the invariant tori of an integrable system. Even though the geometric structure of orbits in phase space changes radically as we move away from integrability, allowing chaotic orbits and substructures of foliated tori around closed orbits, the transition is smooth dynamically, i.e., motion is still mostly quasitoroidal within a given resolution [10].

The solution of the inverse problem is based on the large number N of observed (x, v), from which the distribution function $f(x, v)$ can be estimated. With smaller effective number of phase-space dimensions than six, i.e., in the case of potential symmetries, N may be sufficient for actual binning in sufficiently small volumes $\Delta x \Delta v$ in the reduced phase space. Thus we sample $f(x, v)$ directly by counting and can take it to be our observable, as in section 4. For six dimensions, we can, e.g., take random samples of marginal distributions to be our observables, as will be discussed in section 5.

3.2. Self-consistency regularization. If we assume $\tilde{f}$ to represent all matter in the system, the self-consistency requirement of $\tilde{f}$ and Φ can be used as regularization. For example, we can minimize the discrepancy between (8) and (9) at each space volume with the function

$$\lambda \int |\rho_D(x)| \, d^3x,$$

where λ is a suitable weight factor, and $\rho_D(x)$ is given by (13). In practice, this can be replaced by

$$(14) \qquad \lambda \sum_i [\rho_D(x_i)]^2,$$

for a large number of test locations x_i, written in a manifest χ^2-form. As will be discussed below, it is sometimes better to compare potential values rather than densities at test points due to numerical instabilities in evaluating Poisson's equation directly.

If any amount of dark matter is allowed, no regularization such as (14) can be used if there is no information on the masses of the stars, i.e., observations correspond to the number density, and the estimated mass would only be based on the regularization. This is easy to see by considering the case of constant density fraction of luminous matter: $\rho_{\text{lum}}(x)/(\rho_D + \rho_{\text{lum}}) = const < 1$ everywhere in $\mathbb{R}^3_x$. Regularization of the form (14) would then always yield exactly $\rho_D = 0$, regardless of the true density ratio. If masses are included in observations, i.e., we would solve for $\tilde{f}$ even without regularization, (14) can be employed, though minimizing the amount of dark matter is not necessarily an appropriate principle then.

4. Examples in spherical symmetry

Spherically symmetric systems have the advantage of being integrable, as now the energy $E(x,v)$ as well as the three components of the angular momentum $P(x,v) = x \wedge v$ are isolating integrals; the fact that there are four integrals instead of three is reflected in that each orbit is constrained to a plane (e.g., [2]). The original Jeans' theorem applies only to systems without such degeneracies, but it can be extended to the spherical case [11], stating that any steady-state distribution function in a potential $\Phi(r)$, where $r = \|x\|$ is the radius, must be of the form $f(E, P)$. Further, since the system is spherically symmetric, the direction of P is uniformly distributed over the unit sphere S^2, so we are only interested in functions of the form $f(E, L)$, where

$$(15) \qquad L(x,v) = \|x \wedge v\|.$$

It is customary (e.g., [2]) to use the concepts of relative potential and energy Ψ and $\mathcal{E}$ such that

$$(16) \qquad \Psi(x) := -\Phi(x) + \Phi_0, \qquad \mathcal{E}(x,v) := \Psi(x) - \frac{1}{2}\langle v,v \rangle = -E(x,v) + \Phi_0,$$

where Φ_0 is chosen such that $f > 0$ for $\mathcal{E} > 0$ and $f = 0$ for $\mathcal{E} \leq 0$ (usually, $\Phi_0 = 0$). We now study potentials $\Psi(r)$ and distribution functions of the form $f(\mathcal{E}, L)$. In this section, we simplify the notation by using $f(\mathcal{E}, L)$ directly to denote the mass distribution function (i.e., $\tilde{f}$) as it is the form the analytical treatment here always refers to. For simplicity, let us also assume all the stars to have the same mass m. The systems as well as inversion methods used in this section are simple ones, as

the goal is to investigate the basic properties of the inverse problem rather than develop a full-fledged analysis machinery.

Our phase space is now essentially reduced to radius r and velocity v in a plane. Because of this reduction in variables, spherically symmetric cases offer analytical tractability to a much larger extent than other systems. Of course spherical symmetry rarely occurs in actual gravitational systems, but it is highly useful for exploring the basic properties of the inverse problem (uniqueness, stability, the effect of sampling density, convergence of solution, etc.) in practical computation. Indeed, because of their analytical prospects, spherical systems and distribution functions of the form $f(\mathcal{E}, L)$ have enjoyed ever-continuing popularity in theoretical dynamics for over a century.

Since now we know the exact functional forms of the isolating integrals, we can immediately see that the $\gamma(x)$-uniqueness of remark 7 holds exactly for any nontrivial $f(\mathcal{E}, L)$ and $\Psi(r)$. All distribution functions $f(x, v)$ now have to be of the form $f[\mathcal{E}(x, v), L(x, v)]$, and obviously no function $f[\mathcal{E}(x, v), L(x, v)]$ is separable in the form $h(x)g[\mathcal{E}(x, v), L(x, v)]$. Thus there is no bias function $\xi(x) \neq \gamma(x)$ allowing a product $\xi(x) f'_\psi(x, v)$ to be identical with $\gamma(x) f_\Psi(x, v)$ everywhere (note that the bias function γ does not have to be spherically symmetric). Thus we can state the following uniqueness theorem:

Theorem 4.1. *If a bias function $\gamma(x)$ depends on x only: $\gamma : \mathbb{R}^3_x \to \mathbb{R}$, the product $\gamma(x)f(x, v)$, defined everywhere in $\mathbb{R}^3 \times \mathbb{R}^3$ in a spherically symmetric steady-state system, uniquely determines $\gamma(x)$.*

Remark 8. Similarly, the restricted form $f[\mathcal{E}(x, v), L(x, v)]$ corroborates remark 5 algebraically.

Remark 9. Stäckel potentials $\Upsilon(x)$ [2, 4, 5] are a class of integrable potentials for which the Hamilton-Jacobi equation is separable in ellipsoidal coordinates. Just like in the spherical case, the functional forms of their isolating integrals in the distribution function $f[E(x, v), I_2(x, v), I_3(x, v)]$ are known exactly [4]. Examining the forms of the isolating Stäckel integrals, we can immediately see that corresponding uniqueness results hold for all steady-state systems governed by Stäckel potentials $\Upsilon(x)$.

The simplest nontrivial distribution functions are obtained by just removing L-dependence and using isotropic $f(\mathcal{E})$. The solution for a spherically symmetric system whose $f(\mathcal{E})$ is self-consistent with its density $\rho(r)$ and the corresponding potential $\Psi(r)$ is now simple to obtain. By writing (8) in spherical symmetry and changing the integration variable $v \to \mathcal{E}$, we obtain an Abel integral equation for $f(\mathcal{E})$, the solution of which is the well-known Eddington's formula [2]

$$(17) \qquad f(\mathcal{E}) = \frac{1}{\sqrt{8}\pi^2} \Big[\int_0^{\mathcal{E}} \frac{d^2\rho}{d\Psi^2} \frac{d\Psi}{\sqrt{\mathcal{E} - \Psi}} + \frac{1}{\sqrt{\mathcal{E}}} \Big(\frac{d\rho}{d\Psi}\Big)_{\Psi=0} \Big],$$

where Ψ is used as the argument of ρ via $\Psi(r) \to r(\Psi)$; see [2] for the conditions for $f(\mathcal{E})$ to be non-negative, etc. Note that (17) can naturally be used for non-self-consistent distribution functions (systems containing dark matter) as well. For example, we can solve for the distribution function $f_{\text{lum}}(\mathcal{E})$ of luminous matter from the luminous fraction of the total density: $\rho_{\text{lum}}(r) = \varrho(r)\rho(r)$, with a given $0 < \varrho(r) < 1$, thus replacing $\rho \to \varrho\rho$ in (17) while retaining the $\Psi(r)$ corresponding to $\rho(r)$.

With (17), we can simulate observations of the distribution function of a spherically symmetric system, and then numerically check how well its potential can be recovered. If self-consistency is used for regularization, Poisson's equation (9) is simple to integrate, so the regularizing function corresponding to (14), but in terms of potential rather than density, can be given as

$$
\begin{aligned}
(18) \quad \chi^2_{\text{reg}} = \sum_i \Big(& \Psi(r_i) - 16\pi^2 G[\frac{1}{r_i} \int_0^{r_i} \int_0^{\sqrt{2\Psi(r')}} f[\mathcal{E}(\Psi(r'), u)] u^2 du r'^2 dr' \\
& + \int_{r_i}^{\infty} \int_0^{\sqrt{2\Psi(r')}} f[\mathcal{E}(\Psi(r'), u)] u^2 du r'^2 dr'] \Big)^2
\end{aligned}
$$

where $u := \|v\|$. When using gradient-based methods in optimization, integrals of the form (18) must be differentiated algorithmically for consistency, i.e., by taking the derivatives of the numerical algorithm for evaluating (18) rather than first taking the analytical derivative of (18) and then evaluating the integral numerically.

The lemma (remark 2) and theorem 4.1 ensure that there is a unique solution for both the potential $\Psi(r)$ and the bias function $\gamma(x)$. The χ^2-function of our inverse problem can now be represented as

$$
(19) \qquad \chi^2 = \sum_i \left[N_i m_i - \int_{\Omega_i} \gamma(x) f[\mathcal{E}(\Psi(r), u)] \, d^3x \, d^3v \right]^2 + \lambda \chi^2_{\text{reg}},
$$

where N_i is the number of stars observed in the region Ω_i, and $m_i = m$. To ensure positivity, let us model our potential and distribution function as

$$
(20) \qquad \Psi(r) = \exp\left(\sum_i a_i r^i \right), \qquad f(\mathcal{E}) = \exp\left(\sum_i b_i \mathcal{E}^i \right).
$$

We could include an explicit $1/\mathcal{E}$-term in the exponential sum for $f(\mathcal{E})$ to make sure that $f \to 0$ as $\mathcal{E} \to 0$, but this turns out to be neither necessary nor useful in practice. Because of the adopted form of $\Psi(r)$, it is especially useful to employ (18) rather than density regularization as Poisson's equation would introduce a singularity for one term.

Simple generalizations of the r^{-1}-type Kepler potential of a point mass already yield interesting examples of systems of continuous mass distribution. In fact, dynamical systems whose potentials decrease much faster than this exhibit instability problems [2]. This is another reason why a number of variations of the r^{-1}-theme have been developed. For example, the isochrone potential

$$
(21) \qquad \qquad \Psi_{IC}(r) = \frac{GM}{b + \sqrt{b^2 + r^2}},
$$

where M is the mass of the system and $b > 0$ a scaling constant, with its density pair $\rho_{IC}(r)$ given by Poisson's equation, yields the distribution function [2]

$$
\begin{aligned}
(22) \quad f_{IC}(\mathcal{E}) = \frac{M}{(32GMb)^{3/2}} \frac{\sqrt{\tilde{\mathcal{E}}}}{(1 - \tilde{\mathcal{E}})^4} & \Big[27 - 66\tilde{\mathcal{E}} + 320\tilde{\mathcal{E}}^2 - 240\tilde{\mathcal{E}}^3 \\
& + 64\tilde{\mathcal{E}}^4 + 3(16\tilde{\mathcal{E}}^2 + 28\tilde{\mathcal{E}} - 9)\frac{\arcsin\sqrt{\tilde{\mathcal{E}}}}{\sqrt{\tilde{\mathcal{E}}(1 - \tilde{\mathcal{E}})}} \Big],
\end{aligned}
$$

where $\tilde{\mathcal{E}} = \mathcal{E}b/(GM)$. We will use this in our numerical examples below.
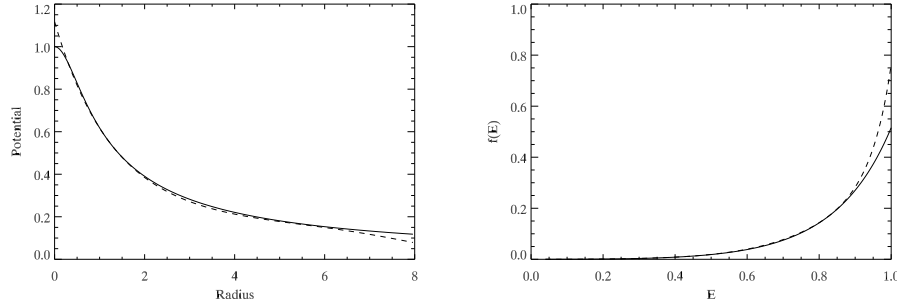
FIGURE 1. (a) $\Psi(r)$ and (b) $f(\mathcal{E})$ of the isochrone potential (solid lines), together with the best-fit models from observed binned $f(x, v)$ (dashed lines).

We simulated observed N_i by creating 400 randomly distributed Ω_i in (r, u)-space (the extents of each Ω_i about 1% of the maximum widths in r and u, with a suitable cutoff value $r_{\max} \gg b$) and computing the amount of mass in them from f_{IC}. Any distribution function corresponding to an r^{-1}-type potential drops fast as r increases; thus those Ω_i close to $r = 0$ contain most of the mass and dominate χ^2 if the volumes of Ω_i are equal. A more sophisticated observation sampling could use increased volumes away from $r = 0$, but in any case this sampling is artificial (as is the symmetry setup) and designed rather to probe the numerical properties of the problem than to simulate realistic conditions.

The objective function (19) was minimized with the Levenberg-Marquardt gradient-based procedure [13]. The initial guess for $\Psi(r)$ was based on the top velocities at each r, and the initial $f(\mathcal{E})$ simply had a correspondingly scaled zeroth-order term and a first-order term anticipating $f \to 0$ as $\mathcal{E} \to 0$. Even with this crude initial values, the procedure converged well towards the correct minimum, i.e., χ^2 does not have many local minima far away from each other in the parameter space at least with reasonably low truncation degrees of (20). The minimum is not particularly sharp, i.e., there are many virtually as good solutions (χ^2-level varying a few percent) close to each other even in the noiseless case, as expected because of the insufficiency of the model ("model noise"), in particular near the centre $r = 0$.

The solution was stable to typical error levels of data: adding 5% or even 10% noise to the data values resulted in virtually the same $\Psi(r)$ and $f(\mathcal{E})$. An exponential bias function $\gamma(x) = \exp(-\|x - x_0\|/R_0)$ with the centre at different values of x_0 was well solved for by the procedure, together with $\Psi(r)$ and $f(\mathcal{E})$. Fig. 1 a and b show the solution $\Psi(r)$ and $f(\mathcal{E})$ (dashed lines), obtained with coefficients up to degrees 4 and 5, respectively, against the actual Ψ_{IC} and f_{IC} (solid lines) with $GM = 1$ and $b = 0.5$ ($r_{\max} = 10$), a noise level of 5%, and a bias $\gamma(x)$ with $R_0 = 5$ and $\|x_0\| = 2$. With the bias, the sample cells in $\mathbb{R}^3_x$ were defined by r and the polar angle θ (x_0 was placed at $r = 2$ and $\theta = 0$).

Including the regularization term (18) for a self-consistent case improved the $\Psi(r)$ fit at large r; as a counterexample, Fig. 1 a shows the case without regularization, corresponding to allowing dark matter; with strong regularization, the lines essentially coincide at all $r < r_{\max}$. The slight "overshooting" of $\Psi(r)$- and $f(\mathcal{E})$-models
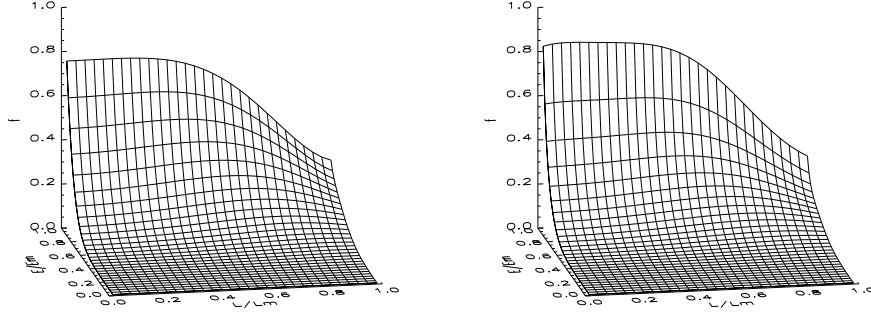
Figure 2. (a) A chosen $f(\mathcal{E}, L)$ for the isochrone, and (b) the best-fit model. The plots are in $(\mathcal{E}/\mathcal{E}_m, L/L_m(\mathcal{E}))$-space.

at the centre of the system (at $r = 0$ and $\mathcal{E} = \Psi(0)$) is partly an intrinsic limitation of the polynomial model.

The form $f(\mathcal{E}, L)$ for the distribution function describes the anisotropy of velocities; there are now infinitely many $f(\mathcal{E}, L)$ corresponding to a given density $\rho(r)$. Creating general functions of this form even for spherical symmetries is not as straightforward as Eddington's formula for $f(\mathcal{E})$. The Osipkov-Merritt procedure [2] is a standard technique for creating a family of self-consistent distribution functions with an adjustable anisotropy parameter, but it is heavily restricted as it still essentially uses Eddington's formula by fixing a combination $\mathcal{U} = \mathcal{E} - cL^2$ and stating $f = f(\mathcal{U})$.

In our case we can (and must) allow the existence of dark matter, so we can carry out the simulations without the self-consistency requirement. Thus we can simply use, e.g.,

$$(23) \qquad\qquad f(\mathcal{E}, L) = f_0(\mathcal{E})h(\mathcal{E}, L),$$

where $f_0(\mathcal{E})$ is any suitable isotropic function, and the function $0 \le h(\mathcal{E}, L) \le 1$ can be chosen to represent any desired dynamical characteristics of the observed matter (anisotropy of angular momentum distribution in various regions, etc.) via some adjustable parameters. The density $\rho(r)$ from this is always $\rho(r) \le \rho_0(r)$, and the difference $\rho_0 - \rho$ is attributed to dark matter; whether or not this interpretation is realistic is not relevant to this study.

Suppose we want to describe a system with central orbits mostly radial and those reaching high radii mostly tangential. Then we can choose, for example,

$$(24) \qquad h(\mathcal{E}, L) = \frac{1}{1+d}\Big\{1 - d\cos\Big[\pi\Big(\frac{L^2}{L_m^2(\mathcal{E})} + \frac{\mathcal{E}}{\mathcal{E}_m}\Big)\Big]\Big\},$$

where $0 \le d \le 1$ is an amplitude factor, $\mathcal{E}_m$ is the maximal $\mathcal{E}$, and the maximal angular momentum $L_m = r\, u(r, \mathcal{E})$ occurs for given $\mathcal{E}$ at $\hat{r}$ for which $dL_m(\hat{r})/dr = 0$, i.e., $L_m(\mathcal{E})$ is obtained from

$$(25) \qquad 2\mathcal{E} = \hat{r}\frac{d\Psi_0(\hat{r})}{dr} + 2\Psi_0(\hat{r}) \Rightarrow \hat{r}, \qquad L_m(\mathcal{E}) = \hat{r}\sqrt{2(\Psi_0(\hat{r}) - \mathcal{E})},$$

where $\Psi_0(r)$ corresponds to $\rho_0(r)$.

Using the isochrone for $f_0(\mathcal{E}) = f_{IC}(\mathcal{E})$ and setting $G = 1$, we have

$$(26) \qquad L_m^2(\mathcal{E})_{IC} = Mb\Big(\frac{M}{2\mathcal{E}b} + \frac{2\mathcal{E}b}{M} - 2\Big).$$

Fig. 2a shows the (suitably normalized) $f(\mathcal{E}, L)$ from $f_{IC}(\mathcal{E})$ (as with the example of Fig. 1) and (24) with an amplitude $d = 1/2$, plotted in the $(\mathcal{E}/\mathcal{E}_m, L/L_m(\mathcal{E}))$-space for convenience. The plotted surface corresponds to $[0.05, 0.95] \times [0.05, 0.95]$ so that the behaviour of $f(\mathcal{E}, L)$ can well be seen without the shrinking L-ranges due to (26). The inverse solution, based on 4000 bins distributed in (r, v)-space in the same manner as in the example of Fig. 1, and with the same computational procedure for solving the inverse problem, is shown in Fig. 2b. The (r, v)-space is here given by the dimensions (r, v_r, v_ϕ), where $v_r = dr/dt$ is the radial velocity and the tangential velocity is $v_\phi = \sqrt{u^2 - v_r^2}$, i.e., $L = rv_\phi$.

With the dark-matter scenario, we have no self-consistency regularization (and $m = 1$ in the $f(\mathcal{E}, L)$-version of (19)). The model for $f(\mathcal{E}, L)$ was taken to be

$$(27) \qquad f(\mathcal{E}, L) = \exp\Big(\sum_{ij} b_{ij}\, \mathcal{E}^i L^j\Big),$$

with the largest degree of 5 for both $\mathcal{E}^i$ and L^j. The two-dimensional polynomial series may not be the best general representation for $f(\mathcal{E}, L)$ (cf. the intrinsic limitations mentioned with Fig. 1), but in this case it at least appears to lead essentially to as good a convergence and solution as in the one-dimensional polynomial case.

5. Distribution function as probability distribution

To be able to solve the inverse problem in terms of probability distributions and without binned sampling, we need to define how to compare a model distribution and observations, i.e., a sampled realization of some probability distribution. Direct comparison methods of distributions are based on cumulative distributions as the latter are well-defined non-binned observables. However, cumulative distributions are uniquely defined only in one dimension though there have been some attempts at two and three dimensions; see [13] and references therein. This dilemma is solved by noting that the one-dimensional marginal distributions of our $f(x, v)$ exactly correspond to the usual line projections in the projection-slice theorem, Radon transform and related tomographic analysis; see, e.g., [3, 9] and references therein.

In general terms, let $f(x)$, $x \in \mathbb{R}^N$, be a probability distribution, and let R, an $N \times N$ orthogonal rotation matrix, $\det \mathsf{R} = 1$, describe a linear transformation to a new basis in $\mathbb{R}^N$: $w = \mathsf{R}x$. The marginal distribution function $h(z)$ along z, any one of the new coordinate axes denoted by the set $W(\mathsf{R})$, is

$$(28) \qquad h(z) = \int_{W(\mathsf{R})\backslash z} f(\mathsf{R}^{-1}w)\, d^{N-1}w,$$

and its cumulative distribution function $C(z)$ is

$$(29) \qquad C(z) = \int_{z_{\min}}^{z} h(z')\, dz',$$

usually with $z_{\min} \to -\infty$. Since $h(z) = dC(z)/dz$ uniquely defines $h(z)$ from a given $C(z)$, we know from tomographic theory (hence this inverse problem can be called dynamical tomography) that:

The probability distribution $f(x)$, $x \in \mathbb{R}^N$, is uniquely determined by the cumulative distributions $C(z)$ of its marginal distributions $h(z)$ along all line directions

in $\mathbb{R}^N$ *that define the coordinate axis directions of* z. For example, in $\mathbb{R}^2$ the lines are defined by all possible values of the rotation angle $\theta \in S^1$, while in $\mathbb{R}^3$ they are given by all the directions $\omega \in S^2$.

Two one-dimensional distributions can be compared via their cumulative distribution functions. In the case of observations and a model, we denote the observational distribution of K observations at z_i (arranged in ascending order by z) by

$$(30) \qquad\qquad S_K(z) = i, \quad z_i \leq z < z_{i+1},$$

for number density, or, if mass is included in our problem,

$$(31) \qquad\qquad S_K(z) = \sum_{j=1}^{i} m_j, \quad z_i \leq z < z_{i+1},$$

with $S_K(z) = 0$ for $z < z_1$. Now we are using cumulative distributions S_K as our observables. The comparison between $S_K(z)$ and a model $C(z)$ can be done with a number of norms; with the usual L_2 we have χ^2-comparison. Another choice often used is the L_∞-norm giving simply

$$\mathfrak{D} = \max_{-\infty < z < \infty} |S_K(z) - C(z)|,$$

with which one can use the Kolmogorov-Smirnov (K-S) probability $0 \leq P \leq 1$ of matching marginal distributions [14, 13], when both $S_K(z)$ and $C(z)$ are normalized to unity by

$$S_K(z) \rightarrow \frac{S_K(z)}{S_K(z_K)}, \quad C(z) \rightarrow \frac{C(z)}{C(z_{\max})},$$

usually with $z_{\max} \rightarrow \infty$, and one normalizes the obtained f afterwards such that $\int f \, d^N x = K$. This, however, distorts the comparison as the high-z ends are now always matched at unity at the comparison stage. Also, K-S probability P is defined by a series with each term a power of

$$d_{KS} = \exp(-K\mathfrak{D}^2),$$

which means that, for a large number of observations K, the model $C(z)$ must be very close to the observed $S_K(z)$ ($K\mathfrak{D}^2 < 10$) for the formal probability $P(d_{KS})$ to have any computationally useful level above zero. Even if the observations were perfect, the model cannot reach such a fit, so usually $P(d_{KS}) \rightarrow 0$ regardless of the model.

It is thus practicable to use the χ^2 from the comparison pairs at each z_i; if K is very large, the comparison can be done for a smaller set of points in each z via pruning or interpolation as the curves of $S_K(z)$ are now so smooth that this loses virtually no information. Thus we can measure the χ^2-sum of marginal distribution comparisons at each chosen coordinate line direction i and its points z_j:

$$(32) \qquad\qquad \chi^2 = \sum_{ij} [S_K^{(i)}(z_j) - C^{(i)}(z_j)]^2.$$

In principle, we could use an arbitrarily large number of projection lines $z^{(i)}$ as, in contrast to standard tomography, there are no limiting factors for choosing them (except for computation time). However, we can expect that even a very limited number of projection lines is compensated for by the strong a priori information (or assumptions) on $f(x, v)$. As is known from, e.g., limited-angle tomography,

incorporating even simple a priori information is a powerful means of enhancing the solution of the inverse problem [9].

We can, for example, expect that simple rotations in each of the 15 two-dimensional coordinate planes in $\mathbb{R}_x^3 \times \mathbb{R}_v^3$ are sufficient for forming a set of projection lines. Now two rows of the 6×6 coordinate transformation matrix R, defining two choices for z, correspond to the rotation through some angle θ, while for other rows R is an identity matrix. We evaluate each $h_i(z)$ corresponding to the choice of z ordered by i using $w = \mathsf{R}_i(x,v)^T$:

$$(33) \qquad h_i(z) = \int \gamma(\mathsf{R}_i^{-1}w)\, f(\mathsf{R}_i^{-1}w)\, d^5w,$$

and find, via $S_K^{(i)}(z)$ and $C^{(i)}(z)$, the best model parameters of $\Phi(x)$, $f[I_\Phi(x,v)]$, and $\gamma(x)$ minimizing (32).

We can use the χ^2-formalism to examine different parts of $\mathbb{R}_x^3 \times \mathbb{R}_v^3$ with different weights, using $[z_{\min}, z_{\max}]$ other than $]-\infty, \infty[$. For example, we can zoom in on parts far away from the centre and fit their cumulative distributions in separate χ^2-terms – otherwise their effect on χ^2 would be negligible compare to centre parts. Also, with marginal distributions we can even use data lacking some coordinates or having poor accuracy in them. For example, if some observations have only recorded x without v, we can still use them in marginal distributions in x.

While a one-dimensional probability distribution is easy to sample using the inverse function of its cumulative distribution, creating a corresponding multidimensional coordinate transform $x \leftrightarrow y$, $x, y \in \mathbb{R}^N$, such that uniform random sampling in y would automatically create a correct distribution in x is difficult and not always realizable. Thus we have to resort to a selective Monte Carlo sampling algorithm in simulations:

Sprinkler algorithm for Monte Carlo sampling an N-dimensional distribution $f(x)$, $f : \mathbb{R}^N \to \mathbb{R}$.

1. Draw a random point $x \in \Omega$ (uniform distribution in $\mathbb{R}^N$), where Ω is the desired region $\Omega \subset \mathbb{R}^N$.
2. Evaluate the probability

$$p(x) = \int_{\omega(x)} f(x')d^N x',$$

   where $\omega(x) \subset \Omega$ is a finite but small integration region around x, similar (with the same volume) for each drawn x.
3. Draw a random number $0 \le q \le 1$ (uniform distribution for $q \in \mathbb{R}$). The point x is included in the sample set $\mathcal{S}$ if $q \le p(x)/p_m$, where $p_m = \max p(x)$, $x \in \Omega$.
4. Return to 1 until $\mathcal{S}$ is large enough to approximate $f(x)$.

As N increases, the ratio between included and drawn x decreases for peaked distributions, so for $N > 3$ this method gets slower than for low dimensions; in such cases faster Monte Carlo sampling can be carried out with Markov chain based MCMC tehniques [9].

As an example, we created 10^5 random samples of the $f_{IC}(\mathcal{E})$ of the previous examples in the (r,u)-plane in the above manner. The distribution function $\hat{f}(r,u)$ in the (r,u)-plane is given by

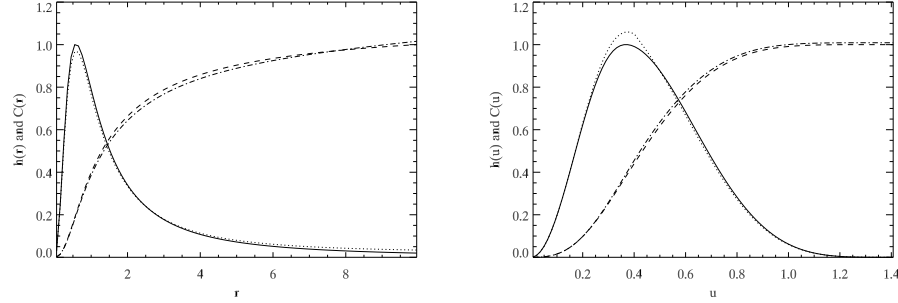$$(34) \qquad \hat{f}(r,u) = 16\pi^2 r^2\, u^2\, f[\mathcal{E}(r,u)],$$

FIGURE 3. Marginal and cumulative distributions in (a) r and (b) u from the isochrone distribution (h in solid line, S_K from samples in dashed line), and the distributions with the best-fit model (h in dotted line, C in dot-dash).

and its marginal distributions in r and u are shown as solid lines in Fig. 3 a and b, suitably normalized to scale between 0 and 1. The cumulative distributions S_K in r and u from the random samples are shown as dashed lines (again suitably normalized). The best-fit model solution C to (32), with $\Psi(r)$ and $f(\mathcal{E})$ of (20) as earlier, matches these well (dot-dash); its marginal distributions are shown as dotted lines. The solution is virtually the same as in Fig. 1, so the tomography of the probability distribution is a practical approach. In particular, just the two distributions $S_K(r)$ and $S_K(u)$ were already sufficient for obtaining the solution: due to the highly restricted form of $f(x,v)$, no more projection lines were needed.

6. CONCLUSIONS AND DISCUSSION

We have defined the inverse problem of dynamical tomography and shown that, with suitable assumptions and mathematical tools, the problem is well-posed and solvable. The uniqueness theorems are central to dynamical tomography: they demonstrate that the steady-state assumption allows a unique solution even with fragmentary data, unobservable mass, and observational bias functions.

Another important concept is the possibility to approximate general (or at least near-integrable) systems with integrable ones. Near-integrable systems that are not very old, but old enough to have settled to a quasistable state, appear closer to integrable than old ones as phase-space diffusion (Arnold diffusion) is not extensive yet. As Nekhoroshev's theorem as well as semianalytical approximations and numerical estimates show [10], the time scale for such diffusion grows fast in inverse proportion to the distance of the initial phase-space point from a KAM torus. For example, in [8] it was shown that even in chaotic zones it is possible to construct approximate tori such that orbits of the system resemble perturbed motion on these tori for small enough time intervals.

Torus construction [6] is a well-defined concept for near-integrable potentials [12]: if a Hamiltonian system is near-integrable in the sense of the KAM theorem, it is integrable on a Cantor set consisting of the surviving KAM tori of the system, i.e., it is possible to construct tori such that their defining action integrals indeed parametrize a global, ordered set. Via torus construction, we find a potential $\Phi(x)$ for which we can create an approximate set of tori (and determine expressions for

their action integrals J and thus for distribution functions $f[J(\Phi(x); x, v)])$ defining an integrable system such that the steady-state integrability assumption used here agrees with the observations as well as possible. Note that $\Phi(x)$ itself does not have to be integrable or even near-integrable; the torus set constructed defines an integrable Hamiltonian that is not usually derivable from a potential, so we never get an approximate integrable potential in the first place. This gives the derived $\Phi(x)$ some additional flexibility; indeed, any integrable potential approximating a real galaxy is probably not a very good representation. The key principle allowing the flexibility is the same as in the uniqueness theorems: we look for isosurfaces in phase space best explaining the observations. The goodness of our solution $\Phi(x)$ is not directly measured by its closeness to an integrable system.

Our $\Phi(x)$ should thus be able to mimic the real potential quite well in cases of near-integrable or even somewhat chaotic but not very old systems. The feasible potentials $\Phi(x)$ that can reproduce, via the torus-construction principle, the approximate isosurface structure of the observed matter distribution can all be expected to be close to each other. In fact, even if $\Phi(x)$ were integrable, we would have to use torus construction to find it even if we started with an integrable initial $\Phi_0(x)$, since the iteration procedure (or equivalent) for fitting a model to observations will generally explore non-integrable potentials $\Phi(x)$.

Let us denote by H_0 the integrable Hamiltonian corresponding to the tori constructed for $\Phi(x)$, and by H the Hamiltonian of $\Phi(x)$. For optimal H_0, the difference (in some chosen norm)

$$\|\delta H\| = \|H - H_0\|$$

is as small as possible over the whole phase space; we call the corresponding tori the optimal tori of Φ, and H_0 the optimal Hamiltonian of Φ. If Φ is integrable, then its optimal tori are its invariant tori: δH vanishes everywhere. For our purposes, $\|\delta H\|$ does not have to be small (although the smaller it is, the better).

We surmise that:

i) The optimal tori constructed for $\Phi(x)$ form a map $\Phi \to H_0$. With the same construction scheme, a potential $\Phi(x) + \epsilon\phi(x)$ yields an optimal Hamiltonian $H_0 + \epsilon h_0$.

ii) The isosurfaces of the distribution function of the system can be approximated by constructing them from a set of 3-tori describable by an integrable Hamiltonian H_0 (but not necesssarily by an integrable potential).

Then our Φ can be expected to be a good approximation of the real potential of the system (as a regularizing constraint, we can choose, e.g., the smallness of $\|\delta H\|$).

The possibility of constructing optimal tori and Hamiltonians holds another great advantage: a multitude of deviations from integrability can readily be modelled with a suitably tailored version of Hamiltonian perturbation theory [7]. This includes both structural detail (resonant orbit families and other details) and timelike irregularities (deviations from steady state).

In a forthcoming study, we will study more general and realistic systems such as axisymmetric and fully three-dimensional Stäckel potentials. We will also introduce more general observational biases and other factors, and investigate their influence. The final goal is the combination of a general torus-construction machinery and an analysis procedure for dynamical tomography, so that we can work with any potentials. With such tools, we can analyze the data from large-scale surveys and construct a consistent mathematical model of the dynamics of our galaxy. The real

galactic problem can be expected to suffer from considerable model noise, so we should employ various forms of modelling the distribution functions and potential. In addition to mapping the density distribution of the dark matter in our galaxy, we should also be able to test whether distributions with alternative theories of gravity are distinguishable from standard models with dark matter.

Acknowledgements

References

[1] V.I. Arnold, "Mathematical Methods of Classical Mechanics" (second edition), Springer, New York 1989.
[2] J. Binney and S. Tremaine, "Galactic Dynamics" (second edition), Princeton 2008.
[3] R.N. Bracewell, *Numerical transforms*, Science, **248** (1990), 607-704.
[4] T. de Zeeuw, *Elliptical galaxies with separable potentials*, Mon. Not. Roy. Ast. Soc., **216** (1985), 273-334.
[5] H. Goldstein, "Classical Mechanics" (second edition), Addison-Wesley, Reading, Mass., 1980.
[6] M. Kaasalainen and J. Binney, *Construction of invariant tori and integrable Hamiltonians*, Phys. Rev. Lett., **73** (1994), 2377-2381.
[7] M. Kaasalainen, *Hamiltonian perturbation theory for numerically constructed phase-space tori*, Mon. Not. Roy. Ast. Soc., **268** (1994), 1041-1050.
[8] M. Kaasalainen, *Construction of invariant tori in chaotic regions*, Phys. Rev. E., **52** (1995), 1193-1196
[9] J. Kaipio and E. Somersalo, "Statistical and Computational Inverse Problems", Springer, New York 2005.
[10] A.J. Lichtenberg and M.A. Lieberman, "Regular and Stochastic Dynamics", Springer, New York 1994.
[11] D. Lynden-Bell, *Only isolating integrals should be used in Jeans' theorem*, Mon. Not. Roy. Ast. Soc., **124** (1962), 1-9.
[12] J. Pöschel, *Integrability of Hamiltonian systems on Cantor sets*, Comm. Pure Appl. Math., **35** (1982), 653-696.
[13] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, "Numerical Recipes" (third edition), Cambridge U. Press, New York 2007.
[14] R. von Mises, "Mathematical Theory of Probability and Statistics", Academic Press, New York 1964.